One- and Multidimensional Conformational Free Energy Simulations

KRZYSZTOF KUCZERA*

Departments of Chemistry and Biochemistry, University of Kansas, 2010 Malott Hall, Lawrence, Kansas 66045

Received 17 August 1995; accepted 26 January 1996

ABSTRACT

A new thermodynamic integration approach to conformational free energy simulations is presented. The method is applicable both to one-dimensional cases (reaction coordinates) and multidimensional situations (free energy surfaces). Analysis of the properties of the thermodynamic integration algorithm is used to formulate methods of calculating multidimensional free energy gradients. The method is applied to calculate the free energy profile for rotation around the central C—C bond of *n*-butane in the gas and liquid phase and to generate maps of the 18-dimensional free energy gradient with respect to all nine ϕ and nine ψ dihedrals of the decaalanine and deca-α-methylalanine peptides in vacuum. For *n*-butane essentially no change in the *gauche–trans* equilibrium between the gas and liquid is predicted within the CHARMM explicit hydrogen model, with the thermodynamic integration, thermodynamic perturbation, and direct simulation methods yielding free energy profiles that are identical within errors. For the decapeptides the right-handed helical region of conformational space is investigated. For decaalanine a minimum on the free energy surface is found in the vicinity of $(\phi, \psi) = (-64.5°, -42.5°)$ in the α-helix region; no minimum exists for 3_{10}-helix-type conformers. For deca-α-methylalanine free energy minima corresponding to both the α-helix at $(-55.5°, -51.5°)$ and the 3_{10}-helix at $(-54°, -29°)$ are found; the α-helix state is favored by about 4 kcal/mol and the barrier for the concerted 3_{10}-helix → α-helix transition is about 3 kcal/mol. The α-methylation also considerably increases the rigidity of the α-helix with respect to deformations. The computational efficiency, ease of generalization to calculations of multidimensional gradients, and analytical capability due to component analysis of free energy differences make the method a novel, powerful tool to improve the basic understanding of conformational equilibria of flexible molecules in condensed phases. A related scheme for energy minimization in the presence of holonomic constraints is also presented, allowing generation of adiabatic energy surfaces in constrained systems. © 1996 by John Wiley & Sons, Inc.

* Author to whom all correspondence should be addressed (e-mail: kuczera@tedybr.chem.ukans.edu).

Journal of Computational Chemistry, Vol. 17, No. 15, 1726–1749 (1996)
© 1996 by John Wiley & Sons, Inc.
CCC 0192-8651 / 96 / 151726-24

Introduction

Free energy simulations have emerged as a powerful tool in physical, chemical, and biological studies of molecular systems.[1-3] Calculating free energy changes is especially valuable, because they are basic observable thermodynamic quantities, related to equilibrium constants and rate constants of processes.

Free energy simulations may be roughly divided into two categories, involving chemical or conformational change. In the latter class, processes involving changes of chemical structure are considered, also known as "computer alchemy."[3,4] Past applications include calculations of differences of solvation energies between different amino acids, nucleic acid bases, small organic molecules and organic and inorganic ions,[5-8] studies of host–guest complexes,[9] chemical reactions,[10,11] design of enzymes and inhibitors,[12] as well as studies of influence of point mutations on various properties of proteins such as substrate binding, thermal stability, quaternary structural changes, and aggregation.[4,13-15]

Conformational changes in molecular systems occur without modifications of chemical structure, and include rotational isomerism, molecular association, and a variety of other processes. Conformational free energy simulations mainly employ three approaches—direct sampling, umbrella sampling, and thermodynamic perturbation theory.[1-3]

In the simplest direct sampling approach, frequencies of occurrence of different values of a conformational coordinate determined from a standard molecular dynamics (MD) or Monte Carlo simulation are converted to a potential of mean force. The *trans–gauche* equilibrium in butane has been studied in this way (see Refs. 2, 3, and 16 for reviews and recent applications). The direct approach fails if the different important conformational states of the studied system are not explored during the time scale of the simulation (e.g., if they are separated by significant barriers).

In the umbrella sampling method an auxiliary potential is added to the Hamiltonian, biasing the system toward a selected region of conformational space which might not be accessible to direct sampling.[2,17] Probability distributions for conformational coordinates are then calculated from molecular dynamics or Monte Carlo simulations. After correcting for the bias introduced by the auxiliary potential, potentials of mean force may be derived from probability distributions. Umbrella sampling calculations have been successfully applied to a wide range of phenomena—from the hydrophobic effect to diffusion of dioxygen through myoglobin (see Ref. 2 for a review).

The thermodynamic perturbation (TP) method proceeds through defining a hybrid potential energy function $U(q, \lambda)$, where q are the system coordinates and λ is called the coupling parameter. $U(q, \lambda)$ has the property of smoothly varying from the initial state $U_0(q)$ to the final state $U_1(q)$ as the coupling parameter λ changes from 0 to 1. When changes to the kinetic energy are neglected, the free energy differences between the states corresponding to coupling parameter values λ' and λ may be evaluated using the "thermodynamic perturbation" formula[1,2,18]:

$$\Delta A(\lambda \rightarrow \lambda') = -kT \ln \langle e^{-\beta[U(q,\lambda')-U(q,\lambda)]} \rangle_\lambda \quad (1)$$

where k is the Boltzmann constant, T the temperature, and $\langle \cdots \rangle_\lambda$ denotes an average over the ensemble corresponding to the system with potential $U(q, \lambda)$. The free energy difference between the initial and final states may then be obtained by accumulating a series of free energy changes spanning the range of λ from 0 to 1. An alternative use of the coupling parameter approach is the thermodynamic integration formula (TI), for the derivative of the free energy with respect to the coupling parameter[1,2,18]:

$$\frac{\partial A(\lambda)}{\partial \lambda} = \left\langle \frac{\partial U(q, \lambda)}{\partial \lambda} \right\rangle_\lambda \quad (2)$$

In this case, the free energy difference between the initial and final states may be obtained by integrating the derivative from eq. (2) over λ from 0 to 1.

The coupling parameter method is applicable to both mutations and conformational change.[1,2] For mutation-type processes, the initial and final state differ by deletion, addition, or substitution of chemical groups. The coupling parameter is an extraneous coordinate and the intermediate potentials $U(q, \lambda)$ usually correspond to unphysical states. For conformational changes, the conformational coordinate of interest, ξ, may be equated with λ; ξ then defines conformations of the molecular system along some path. The important difference between mutational and conformational processes is that in the latter the coupling parameter, or reaction coordinate, is not an external variable, but is part of the specification of system configuration.

The use of both thermodynamic perturbation (TP) and thermodynamic integration (TI) methods has been widespread for mutation type processes.[2,3] For conformational changes, mostly TP-type methods have been implemented so far.[2,19,20] Recently, conformational free energy simulations based on the TI method have been reported in studies of potentials of mean force for side-change rotation in dipeptides in solution[21] and methane association in water.[22] In both TP and TI conformational free energy simulations the same formulae [eqs. (1) and (2)] as for mutational processes have been used, although the standard methods of deriving these equations[1] cannot be applied directly. This is because, in the case of conformational free energy simulations, the coupling parameter is a system coordinate and is involved in the averages in eqs. (1) and (2) in a complicated way.

The focus of this work is on analysis of properties of the thermodynamic integration conformational free energy (CFTI) method. At the start, precise statistic mechanical definitions of conformational free energy, energy, and entropy are given, yielding insight into the physicochemical meaning of these quantities. Generalizations of the standard, one-dimensional thermodynamic integration approach toward calculations of multidimensional conformational free energy gradients and higher derivatives are presented. This framework may be used to develop new formal proofs of the well-known basic thermodynamic integration equations for derivatives of free energy[1] and entropy[22] with respect to a general conformational coordinate. The approach developed here points out a number of theoretical and computational advantages of the CFTI over the TP approach. First, the CFTI method of calculating conformational free energies and entropies is highly computationally efficient—evaluations of multidimensional free energy gradients can be obtained at a cost similar to standard molecular dynamics simulations. Second, the straightforward generalization to multidimensional problems opens the door for conformational analysis on the free energy surface —determination of free energy minima, transition states, and minimum free energy paths for conformational transitions. Third, it is possible to perform decompositions of CFTI free energy profiles into contributions from different energy terms (internal deformation, van der Waals, electrostatic) and parts of the system (solute, solvent, individual chemical groups). Such decompositions have been employed successfully in mutation-type processes to gain insight into the microscopic mechanism of various phenomena involving macromolecules.[4,14,15] Overall, the CFTI approach emerges as a powerful new tool for present and future investigations of conformational equilibria in flexible molecular systems.

A related scheme for energy minimization in the presence of holonomic constraints is also presented, allowing generation of adiabatic energy surfaces in constrained systems.

The theoretical part of the article is presented next; applications of the CFTI method to calculations of the free energy profile for rotation around the central C—C bond in *n*-butane and to free energy gradient maps of decaalanine and deca-α-methylalanine peptides are given in the Simulation Details and Results and Discussion sections. Conclusions are presented last. The basic equation for the multidimensional conformational free energy simulations is proved in Appendix A. The implementation of energy minimization with holonomic constraints is described in Appendix B; the details of the implementation of the CFTI method are presented in the Appendix C; and an analysis of the statistical and systematic errors of the simulations is given in Appendix D.

THEORY

The basic problem of conformational free energy simulations is the determination of the free energy profile $A(\xi)$, also known as the potential of mean force, giving the dependence of the Helmholtz free energy A on a conformational coordinate ξ. Consider a molecular system of N atoms, e.g., a solute in a solvent or a macromolecule, with a separable Hamiltonian $H(p,q) = K(p) + U(q)$, where K is the kinetic and U the potential energy, q are the $3N$ coordinates, and p the $3N$ conjugate momenta. In the canonical ensemble of statistical mechanics, at constant number of particles N, volume V, and temperature T, the probability distribution in coordinate space q is:

$$\rho(q) = \frac{1}{Z} e^{-\beta U(q)}; \qquad Z = \int dq\, e^{-\beta U(q)} \tag{3}$$

with the Z being the classical configurational integral and $\beta = 1/kT$, where k is the Boltzmann constant.

FREE ENERGY PROFILE

A general conformational coordinate ξ—a distance, planar, or dihedral angle, or a more complicated construct—may be defined by specifying a function $\xi(q)$. We can evaluate the probability distribution of ξ in the canonical ensemble[23,24]:

$$\rho(\xi') = \int dq\, \rho(q)\delta(\xi(q) - \xi') = \frac{1}{Z}\int dq\, e^{-\beta U(q)}\delta(\xi(q) - \xi') = \frac{Z(\xi')}{Z} \quad (4)$$

where:

$$Z(\xi') = \int dq\, e^{-\beta U(q)}\delta(\xi(q) - \xi') \quad (5)$$

is the contribution to Z of the "slice" of configuration space with fixed $\xi = \xi'$. The conformational free energy profile may be defined as[19,24-26]:

$$A_c(\xi') = -\beta^{-1}\ln Z(\xi') = -\beta^{-1}\ln \rho(\xi') - \beta^{-1}\ln Z \quad (6)$$

To deal only with logarithms of dimensionless quantities, it is simplest to consider only differences between the free energies of a given conformation ξ' and some reference state ξ_0':

$$\Delta A_c(\xi') = A_c(\xi') - A_c(\xi_0') = -\beta^{-1}\ln\frac{Z(\xi')}{Z(\xi_0')} = -\beta^{-1}\ln\frac{\rho(\xi')}{\rho(\xi_0')} \quad (7)$$

$\Delta A_c(\xi)$ is a well-defined statistical mechanical quantity describing the dependence of the Helmholtz free energy on the conformational coordinate ξ. Because, in practice, one is usually interested only in free energy differences between conformers, using $\Delta A_c(\xi)$ as a measure of conformational free energy does not involve any loss of information. The connection between the conformational free energy and conformer populations $\rho(\xi)$ has been discussed previously.[19,24-26]

As long as the subspaces of fixed ξ form disjoint sets which add up to the whole phase space, we can define the conditional probability distribution $\rho(q/\xi')$, i.e., the probability distribution of configurations q given that the chosen conformational coordinate has a definite value $\xi = \xi'$[27]:

$$\rho(q|\xi') = \frac{\rho(q)}{\rho(\xi')} = \frac{1}{Z(\xi')}e^{-\beta U(q)}\delta(\xi(q) - \xi') \quad (8)$$

and the canonical average of a quantity $X(q)$ over the subspace of configurations with fixed ξ:

$$\langle X\rangle_{\xi'} = \int dq\, X(q)\rho(q|\xi') = \frac{1}{Z(\xi')}\int dq\, X(q)e^{-\beta U(q)}\delta(\xi(q) - \xi') \quad (9)$$

$\langle X\rangle_{\xi'}$ are additive quantities, since we can recover the overall ensemble average by summing up the "slice" contributions.

THERMODYNAMIC INTEGRATION IN ONE DIMENSION

The well-known basic formula of the thermodynamic integration approach gives the derivative of the free energy profile $\Delta A_c(\xi')$ with respect to an arbitrary conformational coordinate ξ in terms of the statistic mechanical average of the corresponding derivative of the potential U over conformations with fixed ξ[1,21,22,24]:

$$\frac{\partial \Delta A_c(\xi')}{\partial \xi'} = \frac{\partial A_c(\xi')}{\partial \xi'} = \langle -\mathcal{F}(\xi')\rangle_{\xi'} \quad (10)$$

$$\mathcal{F}(\xi') = -\frac{\partial U}{\partial \xi'} + \beta^{-1}\frac{\partial \ln|J|}{\partial \xi'} \quad (11)$$

According to eq. (10), in the TI scheme the derivatives of the conformational free energy with respect to the chosen ξ are calculated by performing simulations with fixed values of ξ and evaluating averages of the generalized force $\mathcal{F}$. This force includes a derivative of the potential $\partial U/\partial\xi$ and a term involving the Jacobian (J) of the coordinate transformation between the coordinates q and the generalized coordinates $u = (\eta; \xi)$, due to the generally nonlinear nature of $\xi(q)$ (see Appendix A and Ref. 24). Eq. (10) can be derived as the special case of the general formula eq. (22) given in Appendix A, corresponding to $m = 1$ and $X(q) = 1$.

The standard thermodynamic relation $A = E - TS$, decomposes the Helmholtz free energy A into contributions from the average total energy $E = \langle H(p, q)\rangle$ and the entropy S.[28] For a separable

Hamiltonian it is possible to analogously decompose the conformational free energy profile:

$$\Delta A_c(\xi') = \Delta U(\xi') - T\Delta S_c(\xi') \qquad (12)$$

where $\Delta U(\xi') = \langle U \rangle_{\xi'} - \langle U \rangle_{\xi_0'}$ and $\Delta S_c(\xi') = S_c(\xi') - S_c(\xi_0')$, with ξ_0' being the same reference conformation as used in defining $\Delta A_c(\xi')$ in eq. (7). The above equations provide precise statistical mechanical definitions of the conformational energy profile $\Delta U(\xi')$ and conformational entropy profile $\Delta S_c(\xi')$. From the definitions it follows that $\langle U \rangle_{\xi'}$ is an additive quantity, i.e., the sum of the "slice" contributions gives the total canonical average $\langle U \rangle$, while $S_c(\xi')$ and $A_c(\xi')$ are not additive.

The TI formula for the conformational free energy profile has been given previously[22] and is analogous to that obtained for mutation-type processes[1] [the factor $1/kT$ was omitted in eq. (22) from Ref. 1]:

$$\begin{aligned} T\frac{\partial S_c(\xi)}{\partial \xi} &= \frac{\partial}{\partial \xi}\langle U(q)\rangle_\xi - \frac{\partial A_c(\xi)}{\partial \xi} \\ &= \beta[\langle U(q)\mathcal{F}(\xi')\rangle_{\xi'} \\ &\quad - \langle U(q)\rangle_{\xi'}\langle \mathcal{F}(\xi')\rangle_{\xi'}] \\ &\quad + \beta^{-1}\left\langle \frac{\partial \ln|J|}{\partial \xi'} \right\rangle_{\xi'} \end{aligned} \qquad (13)$$

where again terms involving the Jacobian appear due to the generally nonlinear nature of $\xi(q)$. Eq. (13) can be derived as the special case of the general formula eq. (22) given in Appendix A, corresponding to $m = 1$ and $X(q) = U(q)$.

Within the formalism presented here several advantageous properties of the thermodynamic integration conformational free energy (CFTI) approach become apparent. These are computational efficiency, analytical capability, and generality.

The computational efficiency of the CFTI method is due to its use of local quantities such as the potential energy U and its gradient $\partial U/\partial x_i$, which is usually available at each step of a molecular simulation algorithm. Thus there is very little computational overhead in a CFTI calculation compared to a standard MD simulation, the main additional computation involving evaluation of the coefficients $\partial x_i/\partial \xi$ for eq. (24), which are mostly zero (see Appendices B and C).

The analytical capability of the CFTI approach follows from the linearity of eq. (10) in the potential U. This allows decomposition of the CFTI free energy profiles into contributions from different energy terms (internal deformation, van der Waals, electrostatic) and parts of the system (solute, solvent, individual chemical groups). Such decompositions are not rigorous, but have been employed successfully in mutation-type processes to gain insight into the microscopic mechanism of various phenomena involving macromolecules.[4, 14, 15]

Finally, it is possible to straightforwardly generalize the CFTI algorithm to multidimensional problems (free energy gradients) and evaluations of higher derivatives of the conformational free energy surface, as described below.

THERMODYNAMIC INTEGRATION IN MANY DIMENSIONS

We can introduce the joint probability distribution of a set of conformational coordinates ξ_i, $i = 1, \ldots, m$:

$$\rho(\xi_1', \xi_2', \ldots, \xi_m') = \int dq\, \rho(q) \prod_{i=1}^{m} \delta(\xi_i(q) - \xi_i') \qquad (14)$$

Using eq. (22) from Appendix A with $X(q) = 1$, it is easy to show that the gradient of the free energy surface $A_c(\xi_1', \xi_2', \ldots, \xi_m')$ with respect to the set of coordinates is:

$$\frac{\partial A_c(\xi_1', \xi_2', \ldots, \xi_m')}{\partial \xi_k'} = \langle -\mathcal{F}_k \rangle_{\xi_i', i=1,\ldots,m} \qquad k = 1, \ldots, m \qquad (15)$$

$$\mathcal{F}_k = -\frac{\partial U}{\partial \xi_k'} + \beta^{-1}\frac{\partial \ln|J|}{\partial \xi_k'} \qquad (16)$$

that is, to calculate the derivative $\partial A_c/\partial \xi_k$ of the conformational free energy with respect to the kth coordinate ξ_k of the set, we have to calculate the average of the generalized force $\mathcal{F}_k$ over the set of conformations with all coordinates ξ_i, $i = 1, \ldots, m$ fixed. This result is a significant step forward in simulations of molecular conformational processes. It is a starting point for exploration of multidimensional free energy landscapes of molecular systems by enabling the location and characterization of free energy minima (stable conformers), saddle points (transition states), and minimum free energy paths connecting these points. Calculations of the free energy gradient carry very little computa-

tional overhead compared to evaluating a single derivative, since the most costly terms—the potential energy and its Cartesian gradient—are known in advance in standard computer simulations.

HIGHER DERIVATIVES OF CONFORMATIONAL FREE ENERGY SURFACE

A further generalization of the TI approach is the calculation of second (and higher) derivatives of the conformational free energy. For the one-dimensional case, substituting $X = \mathcal{F}(\xi')$ and $m = 1$ into eq. (22), we obtain:

$$\frac{\partial^2 A_c(\xi')}{\partial \xi'^2} = \frac{\partial}{\partial \xi'} \langle -\mathcal{F}(\xi') \rangle_{\xi'} = \left\langle -\frac{\partial \mathcal{F}(\xi')}{\partial \xi'} \right\rangle_{\xi'} - \beta \left[\langle \mathcal{F}^2(\xi') \rangle_{\xi'} - \langle \mathcal{F}(\xi') \rangle_{\xi'}^2 \right] \quad (17)$$

an analog of which had been derived previously for mutation-type processes.[1] Analogously, substituting $X = \mathcal{F}_l$ into the general eq. (22) the Hessian of the free energy surface described by coordinates ξ_i, $i = 1, \ldots, m$ is given by:

$$\frac{\partial^2 A_c(\xi_1, \xi_2, \ldots, \xi_m)}{\partial \xi_k \partial \xi_l} = \left\langle -\frac{\partial \mathcal{F}_l}{\partial \xi_k} \right\rangle - \beta [\langle \mathcal{F}_k \mathcal{F}_l \rangle - \langle \mathcal{F}_k \rangle \langle \mathcal{F}_l \rangle] \quad (18)$$

where in the last equation $\langle \cdots \rangle$ denotes averaging over configurations with fixed values of all coordinates $\xi_i = \xi_i'$, $i = 1, \ldots, m$. The term $\langle -\partial \mathcal{F}_l / \xi_k \rangle = \partial^2 U / \partial \xi_k \partial \xi_l - \beta^{-1} \partial \ln |J| / \partial \xi_k \partial \xi_l$ contains contributions from both the energy Hessian and the nonlinear coordinate transformation. Knowledge of the second-derivatives can, in principle, enhance the rate of exploration of the free energy surface. However, the practicability of evaluating second derivatives needs to be tested, since their determination involves evaluation and averaging of second derivatives of the potential energy, which is computationally intensive.

Taken together, the computational efficiency, analytical capability and general applicability of the thermodynamic integration conformational free energy simulation approach make it a powerful tool in studies of conformational processes of flexible molecules.

Simulation Details

The systems chosen as the test cases for the new implementation of the CFTI simulation method developed in this work are *n*-butane, as an example of a one-dimensional application and the alanine and α-methylalanine decapeptides as examples of the multidimensional approach.

n-Butane is a simple model system for the general problem of conformational equilibria flexible molecules. The conformational equilibrium of *n*-butane has been widely studied using a range of methods including molecular dynamics and Monte Carlo simulations and integral equations (see Refs. 16, 20, and 29 for recent reviews). The goal of the *n*-butane simulations is twofold. First, we wish to test the new thermodynamic integration conformational free energy (CFTI) algorithm by comparing the results of CFTI, thermodynamic perturbation (TP) and direct sampling from molecular dynamics simulations (MD). Second, once the correctness of the overall CFTI free energy results is established, we wish to employ the decomposition analysis to the *n*-butane system to show the new method's potential for providing physicochemical insight into conformational phenomena. The conditions of the simulations were chosen so as to remain close to the normal boiling point of butane at −0.5°C and 1 atm.[30]

Alanine and α-methylalanine homopeptides serve as models of more general peptide and protein systems. Conformational studies of these models have important implications for our understanding of such phenomena as protein folding, conformational equilibria of flexible peptide drugs, thermodynamic stability of ordered structures of short peptides, and structural preferences of different amino acid residues.[26,31-34] In our simulations we address the specific question of the intrinsic and relative stability of the two helical peptide conformations—the α-helix and the 3_{10}-helix in the alanine and α-methylalanine decapeptides. The decapeptide conformations are described by 18 backbone dihedrals. We map out the full 18-dimensional gradient on a two-dimensional grid of points for which all the ϕ and ψ dihedral angles change in concert. We use this information to describe the free energy surface of these systems in the right-handed helical region—especially the α-helix and 3_{10}-helix states and their interconversion pathway. These results provide an illustration of the new types of information that may be obtained

from the multidimensional CFTI approach. The CFTI results are compared with comparable results obtained by other methods.[32,33]

MOLECULAR MODEL

The new CHARMM (Version 22) all-hydrogen model was used to describe the potential energy of the simulated systems[35,36]; this parameterization is identical to the CHARMM94 set as described in Ref. 37. In this model butane consists of 14 atoms; in molecular dynamics simulations with SHAKE[38] constraints imposed on all C—H bonds, there were 26 conformational degrees of freedom per butane molecule. The two decapeptides studied here are acetylated at the N-terminus and amidated at the C-terminus; their chemical formulae are: ALA10, CH_3CO-$(Ala)_{10}$-$CONH_2$; and AIB10, CH_3CO-$(Aib)_{10}$-$CONH_2$, where AIB is α-methylalanine. The ALA10 system consists of 109 atoms and has 248 degrees of freedom after application of SHAKE constraints to bonds involving hydrogen atoms and fixing all ϕ and ψ backbone dihedrals, while AIB10 consists of 139 atoms and has 318 degrees of freedom under the same conditions.

The final CHARMM parameters used here—the CHARMM94 set described in Ref. 37—are somewhat different from preliminary versions described earlier.[20,39] The main differences are a general overhaul of the internal deformation parameters, including introduction of Urey–Bradley-type terms, and changes in Lennard–Jones parameters describing van der Waals interactions. The atomic partial charges are the same as used previously[20,39]: −0.27 for methyl carbons; −0.18 for methylene carbons, and +0.09 for hydrogens (all charges in atomic units). The parameters and charges for the Aib residue were transferred from the standard Ala. In energy calculations an atom-based 12.0-Å nonbonded cutoff distance was employed, with a switching function between 10.0 and 12.0 Å for van der Waals terms and a shift function at 12.0 Å for electrostatics to eliminate discontinuities due to the cutoff.[35] Full nonbonded interactions were calculated for all atoms separated by three or more chemical bonds.

BUTANE ADIABATIC PROFILE

To gain an understanding of the role of internal flexibility in the conformational equilibrium of butane, the adiabatic rotation profile $\Delta U_m(\phi)$ for isolated *n*-butane was calculated. $\Delta U_m(\phi)$ was obtained from a series of potential energy minimizations with fixed values of ϕ. The adiabatic profile thus gives the lowest possible energy of the flexible *n*-butane molecule for a given value of the central C—C torsion angle. Energy minimizations with fixed values of internal coordinates is performed as described in Appendix B.

BUTANE GAS PHASE FREE ENERGY SIMULATIONS

An isolated butane molecule was heated by random velocity assignments during 10 ps and equilibrated at 270 K for 10 ps in a molecular dynamics simulation. A series of 19 simulations was then performed, in which the value of the central C—C dihedral ϕ of the molecule was fixed consecutively at $\phi = 180, 170, \ldots, 10, 0°$ using the holonomic constraint method of Tobias and Brooks.[40] Each simulation consisted of a 20-ps equilibration and 40-ps trajectory production. The production phase trajectory was thus 19 × 40 ps-760 ps in length and corresponded to an average temperature of 275 ± 50 K. In all simulations the Verlet algorithm was used with a time step of 2 fs and SHAKE[38] constraints imposed on all C—H bonds.

Values of the potential energy U and its derivative with respect to changes in the central C—C dihedral $\partial U/\partial\phi$ from the trajectories were stored every 0.04 ps. Free energy and entropy derivatives were calculated according to eqs. (10) and (13) with $\xi = \phi$. Free energy, energy, and entropy profiles were calculated by integrating the corresponding derivatives, taking the *trans* conformer as the reference state. $\Delta U(\xi)$ was determined using eq. (12). Linear interpolation of the derivatives was used to generate profiles with 5° spacing.

Component analysis was done in the postprocessing phase, by reading the stored molecular coordinates, calculating the energy gradients due to the selected energy terms, and averaging according to eq. (10). The internal deformation energy consists of bond stretching, angle bending, Urey–Bradley terms, and proper and improper dihedral deformation.[35] For an isolated butane molecule only intramolecular nonbonded interactions, consisting of electrostatic and van der Waals interactions between atoms separated by three or more bonds, have to be considered.

Based on the same trajectories, thermodynamic perturbation (TP) calculations of the free energy profile were also performed, using the method of Tobias and Brooks.[20] The TP conformational en-

tropy profile was calculated from the numerical derivative of the partition function with respect to temperature, using a temperature step ΔT = 10 K.[20]

LIQUID BUTANE FREE ENERGY SIMULATIONS

All liquid butane calculations described here were performed for a truncated cubic cell[41] based on a cube of edge 28.437 Å containing 72 butane molecules, at a density of 0.602 g/cm^3, corresponding to liquid butane at −0.5°C and 1 atm.[30] The system was prepared by placing *trans n*-butane molecules on a regular space grid within the cell, energy minimization, 10 ps of heating by random velocity assignments at stepwise increasing temperatures, and 20 ps equilibration at 272 K by velocity rescaling every 0.4 ps. Starting from the final structure of this initial equilibration, a series of 19 simulations was performed, in which the value of the central C—C dihedral ϕ of one of the molecules was fixed consecutively at ϕ = 180, 170, . . . , 10, 0° using the holonomic constraint method of Tobias and Brooks.[40] The remaining 71 butane molecules had unconstrained dihedrals and were able to sample a wide range of conformations during the simulation (see Direct Simulations section). Each free energy simulation consisted of a 20-ps equilibration at 272 K and a 40-ps molecular dynamics trajectory production run. In each case, the final structure from the trajectory production with $\phi = \phi_0$ was used to start the simulation for $\phi = \phi_0 - 10°$. The total production phase trajectory was 19 × 40 ps = 760 ps in length and corresponded to an average temperature of 272 ± 6 K.

Equilibrium constants were calculated as ratios of probabilities of the *trans* state (120°−180°) to the *gauche* state (0°−120°). The probabilities were obtained directly from the free energy profiles.[25]

Components of the TI free energy profile were calculated from the internal deformation and nonbonded interactions, analogously, as in the gas phase simulations. The nonbonded contribution was further divided into intramolecular and intermolecular terms, with the intermolecular part consisting of electrostatic and van der Waals interactions of the selected butane molecule with its neighbors within the simulation cell and from periodic images within the nonbonded cutoff.

The Verlet algorithm was used with a time step of 2 fs and SHAKE[38] constraints imposed on all C—H bonds. This relatively long time step has been found to yield stable trajectories with good energy conservation properties as long as X—H bonds are constrained.[42] All liquid simulations were performed using periodic boundary conditions with a constant number of particles, volume, and total energy (NVE).[41] Using the theorem on equivalence of ensembles,[28] we assume that average values computed under NVE conditions with an average temperature T are equivalent to those from an NVT simulation at the same temperature.

DIRECT SIMULATION OF BUTANE CONFORMATIONAL EQUILIBRIUM

To test the CFTI approach further, the distribution of *n*-butane conformers were evaluated directly from unconstrained molecular dynamics simulations.

In the case of liquid butane, the distributions of the dihedral angle ϕ were obtained for the 71 unconstrained molecules over the 760-ps total production phase trajectory generated for the free energy simulations. The dihedrals were sampled every 0.04 ps for each molecule, giving a set of 1,349,000 data overall for evaluating the probability distribution $P(\phi)$. The distribution was symmetrized by adding up contributions from angles of $\pm\phi$. To enable direct comparison with the free energy simulations the probability distribution was converted to a free energy profile according to eq. (7) with the *trans* conformer ($\phi = \pi$) taken as the reference.

In the gas phase a 150-ns trajectory was generated for an isolated butane with unconstrained dihedrals, corresponding to an average temperature of 280 ± 60 K. By sampling the dihedral ϕ every 0.2 ps a set of 750,000 data readings was generated, yielding a probability distribution $P(\phi)$, which was symmetrized by adding contributions from $\pm\phi$ and converted to a free energy profile $\Delta A_c(\phi)$ as described above.

Equilibrium constants were calculated as ratios of probabilities of the *trans* state (120°−180°) to the *gauche* state (0°−120°).

DECAPEPTIDE FREE ENERGY GRADIENT MAPS

It is unfeasible to perform a systematic exploration of the complete 18-dimensional conformational space of a decapeptide, described by nine ϕ and nine ψ dihedrals. Our simulations explore a limited region of this space, in which with all nine ϕ and nine ψ backbone dihedrals have identical values: $\phi_i = \phi$, $\psi_i = \psi$, $i = 1, \ldots, 9$. The conforma-

TABLE I.

Characterization of Rotational Profile Around Central C — C Dihedral ϕ in *n*-butane. Statistical Errors of the Calculated Quantities are Given, Calculated as Described in Text. The *trans-gauche* Barrier Maximum is at $\phi = 120°$ in the Adiabatic and TP Profiles and at $\phi = 115°$ in TI; the *gauche* Minimum is at $\phi = 65°$ in the Adiabatic and TI Profiles and at $\phi = 67.5°$ in TP.

	Gas				Liquid, 272 K		
			Free energy, 275 K			Free energy	
Feature	Adiabatic ΔU_m	Direct MD 280 K	TI	TP	Direct MD	TI	TP
			Free energy	ΔA_c	kcal / mol		
trans-gauche barrier		3.50 ± 0.01	3.44 ± 0.01	3.49 ± 0.01	3.36 ± 0.03	3.35 ± 0.02	3.44 ± 0.03
Δ(*trans-gauche*)		1.03 ± 0.02	0.96 ± 0.01	0.96 ± 0.02	0.90 ± 0.02	0.95 ± 0.02	1.00 ± 0.04
sgn barrier		6.00 ± 0.19	5.60 ± 0.02	5.72 ± 0.03	5.45 ± 0.16	5.58 ± 0.03	5.70 ± 0.05
			Energy	ΔU	kcal / mol		
trans-gauche barrier	3.49		3.40 ± 0.03	3.46 ± 0.20		3.06 ± 0.80	2.99 ± 2.38
Δ(*trans-gauche*)	0.88		0.95 ± 0.04	0.97 ± 0.36		0.90 ± 1.00	0.86 ± 3.61
syn barrier	5.30		5.50 ± 0.06	5.63 ± 0.55		5.95 ± 1.26	6.00 ± 4.38
			Entropy	$-T\Delta S_c$	kcal / mol		
trans-gauche barrier			0.04 ± 0.03	0.03 ± 0.20		0.29 ± 0.80	0.45 ± 2.38
Δ(*trans-gauche*)			0.01 ± 0.04	−0.01 ± 0.36		0.05 ± 1.00	0.14 ± 3.61
syn barrier			0.10 ± 0.06	0.09 ± 0.55		−0.37 ± 1.26	−0.29 ± 4.38
trans fraction, %	71.8[a]	76.5	77.2	74.5	73.1	78.2	76.3
K_{tg}	0.39[a]	0.31	0.30	0.34	0.37	0.28	0.31

[a] Calculated at $T = 272$ K.

tions of the system may thus be identified by the pair of values (ϕ, ψ). For each of the ALA10 and AIB10 decapeptides, simulations were performed in vacuum, at 117 points on a square grid with 3° spacing, for ϕ varying from $-75°$ to $-51°$ and ψ from $-20°$ to $-56°$. The idealized right-handed α-helix and 3_{10}-helix conformations are expected to fall into this region. At each point the system coordinates were generated from model geometric parameters. A short energy minimizations was followed by a random velocity assignment, a 20-ps equilibration period of molecular dynamics with velocity rescaling to bring the temperature close to 300 K, and a 40-ps trajectory generation phase at 300 K. The ϕ and ψ dihedrals were kept fixed at the required values during the minimization (see Appendix B), equilibration, and dynamics.[20,40] The average temperature of the 117 simulations was 258 ± 45 K for ALA10 and 287 ± 71 K for AIB10.

Since the simulations are performed with all ϕ and ψ fixed, we use eq. (15) to obtain the complete 18-dimensional gradient $(\partial A/\partial\phi_1, \ldots, \partial A/\partial\phi_9, \partial A/\partial\psi_1, \ldots, \partial A/\partial\psi_9)$ at all explored points. To facilitate the analysis of this data the free energy gradient components corresponding to all ϕ and all ψ were added up to give:

$$\frac{\partial A}{\partial \phi} = \sum_{i=1}^{9} \frac{\partial A}{\partial \phi_i}; \qquad \frac{\partial A}{\partial \psi} = \sum_{i=1}^{9} \frac{\partial A}{\partial \psi_i}$$

This procedure leads to an effective two-dimensional free energy surface describing concerted conformational transitions of the peptides. Reducing the dimensionality of the free energy surface from 18 to 2 allows for a more facile analysis of results. By following only concerted changes in the ϕ and ψ we obtain only a partial description of the 18-dimensional conformational space. The availability of the complete 18-dimensional gradient at all explored points enables a more in-depth characterization of the visited parts of the surface.

ERROR ANALYSIS

A discussion of the statistical and systematic errors is given in Appendix D.

APPARENT FORCE TERMS DUE TO JACOBIAN

The apparent force terms involving $\ln|J|$ in eqs. (11) and (16) were neglected in our calculations. Based on evaluations of $\partial \ln|J|/\partial\phi$ for structures along the *n*-butane adiabatic profile, we estimate that the apparent force is of the order of 10^{-6} kcal/(mol rad). This is in accord with previous findings for one-dimensional cases.[24] For ALA10 numerical evaluations of derivatives of $\ln|J|$ with respect to ϕ and ψ for several structures yield apparent forces of the order of 10^{-4} kcal/(mol rad). It is thus clear that for both *n*-butane and the decapeptides the apparent forces can be neglected compared to the contributions from the potential energy derivatives.

PROGRAMS AND COMPUTERS

The molecular dynamics simulations described here were performed using the program CHARMM (Version 22),[35] which has been extended and modified by the author to implement the two new protocols—energy minimization with holonomic constraints and the multidimensional CFTI method—as described in Appendices B and C. The computers used were IBM RS/6000-550 and −375 workstations at the Departments of Chemistry and Biochemistry of the University of Kansas at Lawrence. A 10-ps simulation of the liquid butane system took about 5 hours of CPU time on the Model 375, while one point of the decapeptide vacuum free energy gradient maps, involving 60 ps of molecular dynamics, took about 1 hour of CPU.

Results and Discussion

ADIABATIC PROFILE $\Delta U_m(\phi)$

The adiabatic rotation profile $\Delta U_m(\phi)$ for isolated *n*-butane is presented in Table I and Figure 1. The adiabatic profile gives the lowest possible potential energy of the flexible *n*-butane molecule for a given value of the central C—C torsion angle. This quantity corresponds approximately to the gas phase free energy profiles obtained using simplified models of butane with 1 degree of freedom[16,29,43-45] and is a convenient point of reference in analyzing the free energy profiles and their variation with environment. The CHARMM all-hydrogen potential[35,36] yields a *gauche* energy 0.88 kcal/mol above the *trans* state, and 3.49 and 5.30 kcal/mol for the *gauche–trans* and *syn* barriers, respectively. This is in good agreement with a recent vibrational spectroscopic study, which found the potential energy of the C—C rotation of gaseous *n*-butane having a *gauche–trans* barrier of 3.64 kcal/mol, a *syn* barrier of 5.01 kcal/mol, and a *gauche* energy of 1.10

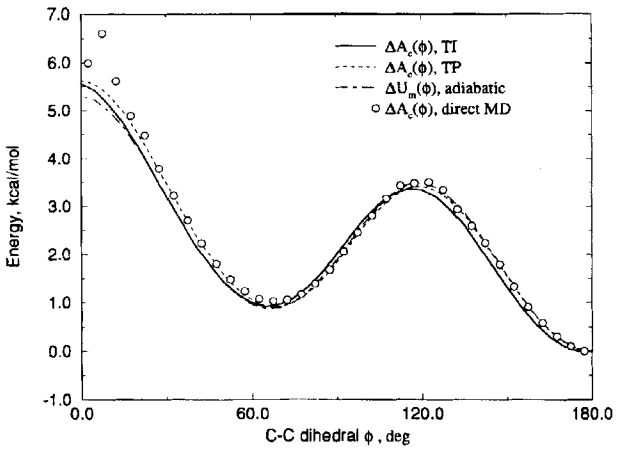

FIGURE 1. Free energy profiles $\Delta A_c(\phi)$ for rotation around the central C — C bond in gaseous *n*-butane. (——): TI free energy simulations, 275 K; (···): TP free energy simulations, 275 K; (·-·-·): adiabatic profile $\Delta U_m(\phi)$; (●): direct molecular dynamics simulation, 280 K.

kcal/mol.[46] Other experimental studies find *gauche* energies of 0.7–1.0 kcal/mol.[47–50] High level quantum chemical calculations place the *gauche* energy at 0.5–1.0 kcal/mol,[51,52] and the two barriers at ca. 4.0 and ca. 8.0 kcal/mol.[52] The previous CHARMM potential gave 0.7 kcal/mol for the *gauche* energy and barrier heights similar to those of the new parameterization.[20,39] The potential used in most previous computer simulations[16,29,30,43–45] with implicit hydrogen atoms yielded an energy level of 0.7 kcal/mol for the *gauche* form and 2.95 and 10.7 kcal/mol for the barriers.[43,53,54]

The *gauche–trans* equilibrium constant calculated from ΔU_m at a temperature of 272 K is 0.39 (see Simulation Details section and Table I). This value is somewhat lower than the 0.5–0.6 found in previous gas phase calculations (see Refs. 20 and 29 for reviews of previous results). The reason for the lower equilibrium constant is clearly the elevation of the adiabatic *gauche* energy in the new CHARMM potential. However, our results are consistent with experimental data which fall mostly in the 0.4–0.5 range (see Refs. 20 and 39).

BUTANE IN GAS PHASE

The free energy profiles $\Delta A_c(\phi)$ for rotation around the central C—C bond of *n*-butane in the gas phase obtained using TI, TP, and direct molecular dynamics (MD) simulations are presented in Figure 1 and Table I.

The TI gas phase free energy calculations at 275 K yield values of 3.44 and 5.60 kcal/mol for the *trans—gauche* and *syn* barriers, respectively. The free energy difference between the *gauche* minimum at $\phi = 65°$ and the *trans* minimum at $\phi = 180°$ was 0.96 kcal/mol. The TI, TP, and direct MD results for $\Delta A_c(\phi)$ are essentially identical, agreeing within several standard deviations (see Table I). The free energy profiles are also quite similar to the adiabatic profile, with the largest difference of about 0.3–0.4 kcal/mol occurring at the *syn* barrier. The direct MD profile deviates from the free energy simulation results in the vicinity of the *syn* barrier. This is due to poor statistics in the unconstrained MD simulation, as only 10 data points out of 750,000 were found between $-15°$ and $+15°$ (see Simulation Details).

The *gauche–trans* equilibrium constants in the gas phase are 0.31 for direct MD (at 280 K), 0.30 for TI, and 0.34 for TP (both at 275 K). All these values are lower than the 0.39 obtained from the adiabatic profile. Our calculations thus indicate that introducing structural flexibility shifts the equilibrium toward the *trans* state. Gas phase equilibrium constants found here are somewhat lower than previous calculations and experimental measurements, which fall mostly in the 0.4–0.6 range (see Adiabatic Profile section). Simulations using an older version CHARMM potential yielded an equilibrium constant of 0.59 in the gas phase.[20]

As could be expected from the similarity of the free energy and adiabatic profiles, the entropic term $-T\Delta S_c(\phi)$ is small along most of the profile, reaching its largest value of about 0.1 kcal/mol at the *syn* maximum. Entropy calculations based on the TP method give similar results. However, the TP entropy values exhibit statistical fluctuations that are about an order of magnitude larger than the TI results (see Error Analysis section). The average potential energy profile $\Delta U(\phi)$ from the TI method gives 0.95 kcal/mol for the *gauche* conformer, and 3.40 and 5.50 kcal/mol for the *trans–gauche* and *syn* barriers. This is in good agreement with a recent vibrational spectroscopic study,[46] although other investigations find *gauche* energies closer to 0.7 kcal/mol (see Adiabatic Profile section).

LIQUID BUTANE

The simulation results for liquid butane are presented in Table I and Figures 2 and 3. The free energy profiles from the TI, TP, and MD calcuations agree within their standard deviations (Table

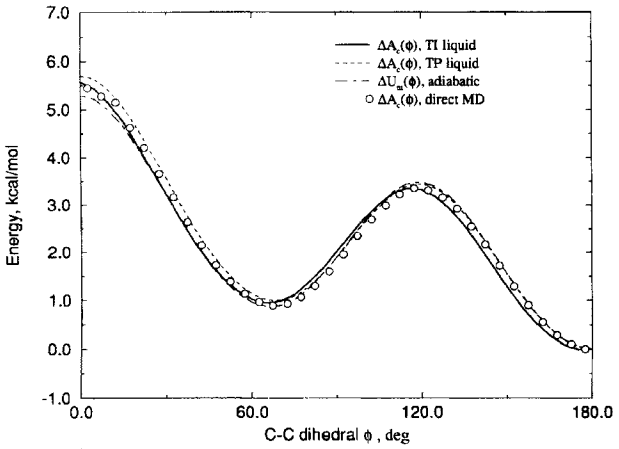

FIGURE 2. Free energy profiles $\Delta A_c(\phi)$ for rotation around the central C — C bond in liquid *n*-butane at 272 K. (——): TI free energy simulations; (· · ·): TP free energy simulations; (·-·-·): gas phase adiabatic profile $\Delta U_m(\phi)$; (●): direct molecular dynamics simulation.

I, Fig. 2). The liquid free energy profiles are also quite similar to the gas phase results (see Table I). The general agreement between all three simulation methods used in this work indicates that the CFTI approach yields correct conformational free energy profiles.

The liquid phase *gauche–trans* equilibrium constants are 0.28 for TI, 0.31 for TP, and 0.37 for the direct MD method. The equilibrium constants calculated here are lower than those of previous calculations, which range from 0.47 to 0.85.[16,29,30,43-45]

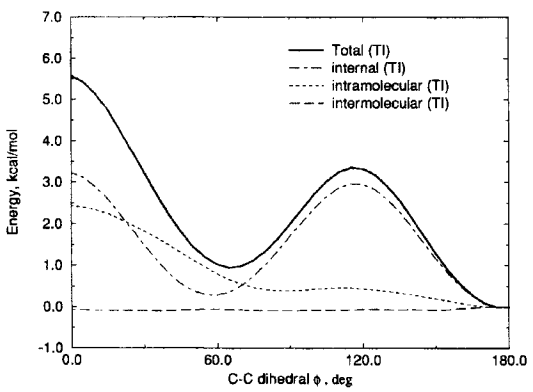

FIGURE 3. Decomposition of the TI free energy profile $\Delta A_c(\phi)$ of liquid butane into components from internal deformations and nonbonded interactions, 272 K. (——): total TI free energy profile $\Delta A_c(\phi)$; (·-·-·): TI internal deformation energy component; (· · ·): TI intramolecular interaction component; (– – –): TI intermolecular interaction component.

The value of 0.47, closest to the results of this work, was obtained in Monte Carlo calculations of Jorgensen.[30] The main reason for these differences appears to be the higher *gauche* energy of the new CHARMM potential. Because the calculated liquid equilibrium constants are indistinguishable from the gas phase values within errors (see Appendix D), the overall conclusion from our simulations is that there is no equilibrium change between *n*-butane in the gaseous and liquid state. The absence of a conformer population shift toward the *gauche* form in the liquid has been found in some previous studies,[30,55,56] whereas significant shifts have been found in others.[16,29,43-45] On the other hand, simulations of butane solutions in water and CCl_4 found a definite increased preference for the *gauche* form compared to gaseous butane.[20,23,37,57] Because a lack of equilibrium shift is present in all of our CFTI, TP, and direct simulation results, it is clearly a property of the molecular model used, which differs in details of description of internal deformations and nonbonded interactions from those employed in other reported butane studies. Several studies indicate a dependence of the *gauche–trans* equilibrium constant on the attractive part of the intermolecular potential.[16,56,58] Our experience shows that another factor hindering the precise calculation of equilibrium shifts is the low precision of the gas phase equilibrium constant for flexible butane models (see Appendix D).

The decomposition of the free energy profile into energetic and entropic terms is presented in Table I. The entropic contribution to the free energy profile is small, and the TI and TP results are in good agreement. The simulations indicate an entropic destabilization of the *trans–gauche* transition state by 0.3–0.4 kcal/mol, a stabilization of the *cis* maximum by a similar amount, and approximately zero contribution to the *gauche* minimum. Since both the maxima showed slight entropic destabilization relative to the *trans* conformer in the gas phase, this difference is a solvent effect. The entropic stabilization of the *syn* transition state may be explained by the decreased volume of that structure which would lead to solvent disorder (see below). The statistical fluctuations of the entropy and average energy are quite large (Table I), thus these results are much less reliable than the free energies.

Figure 3 shows the decomposition of the liquid *n*-butane conformational free energy profile $\Delta A_c(\phi)$ into contributions from internal deformations, and both intra- and intermolecular nonbonded interactions. The interactions between the

unconstrained *n*-butane molecules (the solvent–solvent term) remain unchanged upon infinitesimal perturbation of the central C—C dihedral of the selected butane molecule for which the dihedral is constrained; these terms thus do not contribute to $\Delta A_c(\phi)$. Within their statistical fluctuations, internal deformation and intramolecular interaction components found in the liquid are essentially identical to those found in the gas phase, while the liquid intermolecular component is indistinguishable from zero (see Appendix D). The intramolecular interaction components for angles $\phi > 70°$ are essentially identical to the adiabatic interaction term, while for $\phi < 70°$ the liquid simulation results lie above the adiabatic contribution. This may be explained by conformational crowding in the conformations ranging between *gauche* and *cis*, in accord with chemical intuition. The crowding is illustrated by the decreased molecular volume of these conformations relative to *trans* by about 0.5 $Å^3$ (0.4%), apparently leading to disordering of the neighboring molecules and entropy increase for conformers in the *syn* range.

ALA10 FREE ENERGY GRADIENT MAP

The reduced, two-dimensional free energy gradient map describing concerted transitions of ϕ/ψ dihedrals in the right-handed helical region of the ALA10 peptide (CH_3CO-$(Ala)_{10}$-$CONH_2$) is presented in Figure 4. The arrows in Figure 4 show the direction and rate of increase of the system conformational free energy evaluated on a grid. The CFTI results are used below to explore several features of the ALA10 free energy surface: to identify free energy minima which correspond to stable conformations; calculate free energy profiles along several directions; and decompose free energy changes into contributions from internal strain and nonbonded interactions.

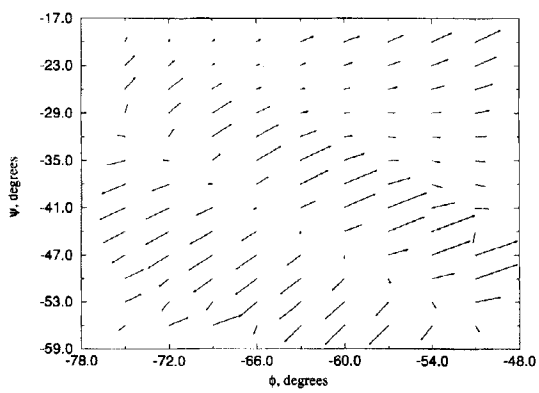

FIGURE 4. The reduced two-dimensional free energy gradient map of the ALA10 peptide. Arrows indicate the direction of increasing conformational free energy ($\partial A_c/\partial\phi, \partial A_c/\partial\psi$). For orientation, the gradient at $(\phi, \psi) = (-60°, -41°)$ is (106.0, 87.0) kcal/(mol rad).

A necessary condition for a free energy minimum is that the reduced gradient ($\partial A_c/\partial\phi, \partial A_c/\partial\psi$) be equal to zero. In agreement with expectations, we find two regions of low gradient in Figure 4, corresponding to the α-helix and 3_{10}-helix structures. The α-helix region is a relatively long trough extending diagonally approximately from $(\phi, \psi) = (-75°, -32°)$ to $(-57°, -47°)$. The conformations found in this region correspond to anticorrelated changes in ϕ and ψ, and exhibit similar $i - i + 4$ hydrogen-bonding patterns. The state with the lowest free energy in the α-helical region of the reduced map lies in the vicinity of $(-66°, -41°)$. This means that conformations in that area are stable with respect to small displacements in which all ϕ and/or all ψ dihedrals are changed in concert, $\Delta\phi_i = \Delta\phi$, $\Delta\psi_i = \Delta\psi$, $i = 1, \ldots, 9$. Analysis of the full 18-dimensional free energy gradient shows that all gradient components change sign from negative to positive as we pass through the $(-66°, -41°)$ point. This indicates that a minimum of the complete 18-dimensional free energy hypersurface exists close to that point and that the α-helix is a stable molecular structure for ALA10 in vacuum.

The 3_{10}-helix region is in the vicinity of $(-60°, -30°)$.[32,33] A corresponding region of low conformational free energy gradient exists in Figure 4. However, the reduced gradient map does not indicate the presence of a free energy minimum in the 3_{10}-helix region—the conformations are stable with respect to small concerted displacements along ψ, but the free energy increases monotonically along ϕ. Inspection of the full 18-dimensional free energy gradient confirms this—most of the derivatives of A_c with respect to the individual ϕ_i retain the same sign in this region. The 3_{10}-helix is thus not a stable structure of ALA10 in vacuum.

Free energy profiles obtained from the ALA10 map are presented in Figures 5 and 6. Figure 5 shows the profile obtained by integrating the free energy gradient along the α-helix trough. The conformation of lowest free energy is close to $(-64.5°, -42.5°)$, in excellent agreement with average values found in α-helices in protein crystals.[59] In this shallow profile, anticorrelated changes of ϕ,ψ by $\pm 4.5°$ lead to free energy changes of only about 0.7 kcal/mol, indicating that the α-helix

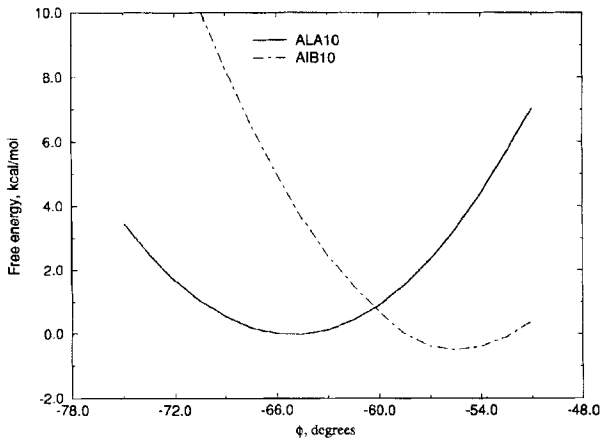

FIGURE 5. Free energy profiles obtained by integration of conformational free energy gradients along the α-helix trough $\phi + \psi = -107°$. (——): ALA10 total free energy; (·-·-·) AIB10 total free energy.

structure can easily be deformed along the trough direction, as has been found in molecular dynamics simulations.[60] The free energy profile in the direction perpendicular to the α-helix trough is quite steep, so that it is difficult to deform the α-helix in this direction; correlated changes of ϕ, ψ by $\pm 4.5°$ lead to free energy changes of about 11.0 kcal/mol.

Figure 6 shows the free energy profile along the $\psi = -60°$ line, which connects the α-helix and 3_{10}-helix regions. Integrating the ALA10 free energy gradient from $(-64.5°, -42.5°)$ to $(-60°, -47°)$ and further along $\phi = -60°$ to $(-60°, -29°)$, we find that within our model the α-helix is more stable than the 3_{10}-helix by 1.37 ± 0.1 kcal/mol for the ALA10 peptide in vacuum. This value is approximate—while our α-helical state is close to the true free energy minimum, the 3_{10}-helix state cannot be uniquely defined since there is no free energy minimum in the 3_{10}-helix region. Our α-helix/3_{10}-helix free energy difference is in agreement with previous simulations, which found a preference for the α-helix conformation in alanine homopeptides.[32,33,61] Simulations of undecaalanine with fixed bonds along the same path as in Figure 6 predicted that the α-helix should be more stable by 13.5 kcal/mol in vacuum, and 10.6 kcal/mol in aqueous solution.[32] The simulations of Jorgensen et al.[32] found a barrier of 4.8 kcal/mol in vacuum and 2.8 kcal/mol in solution between the 3_{10}-helix and α-helix of undecaalanine. Interestingly, we find no barrier between the two helical structures. As can be seen from Figure 4, our data allow generation of profiles connecting the α-helix and 3_{10}-helix with (e.g., along $\phi = -54°$) and without barriers (e.g., along $\phi = -60°$ as in Fig. 6). The advantage of the multidimensional approach is that it provides more information about the nature of the free energy surface than a profile along a single direction.

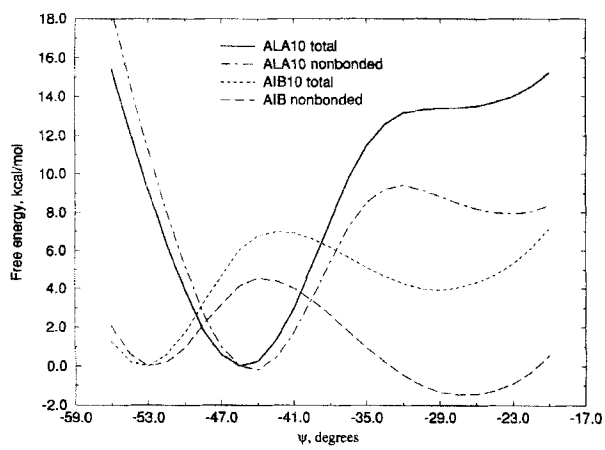

FIGURE 6. Free energy profiles obtained by integration of conformational free energy gradients along paths connecting the α-helix and 3_{10}-helix regions. (——): ALA10 total free energy profile along line $\phi = -60°$; (·-·-·): ALA10 nonbonded interaction component along line $\phi = -60°$; (···): AIB10 total free energy profile along line $\phi = -54°$; (– – –): AIB10 nonbonded interaction component along line $\phi = -54°$.

To gain further insight into the conformational free energy surface of ALA10, the total free energy was decomposed into contributions from internal strain (bond, angle, dihedral deformation) and nonbonded interactions (van der Waals and electrostatic). The internal strain contribution to the gradient was small and approximately constant over the mapped area; the features of the total free energy map—minima and maxima—were determined primarily by the nonbonded terms. Decomposition of the α-helix → 3_{10}-helix free energy profile into internal strain and nonbonded terms is shown in Figure 6. The free energy difference between the α-helix and 3_{10}-helix states is determined primarily by nonbonded interactions. It is interesting to note that the nonbonded component in Figure 6 shows a barrier separating the two helical states for deformations along ψ. Thus, a calculation using the CHARMM parameters and a fixed internal geometry would predict a barrier of ca. 2 kcal/mol between the 3_{10}-helix and α-helix of ALA10. This effect may partly explain the differences between our results and those of Jorgensen et al., who performed simulations with fixed bond lengths.[32]

Energy minimization with fixed ϕ and ψ has been performed for the final structures from the free energy simulations for (−66°, −41°)—the α-helix state, (−60°, −29°)—the approximate 3_{10}-helix state, and (−60°, −35°)—an intermediate, the state where the nonbonded component of Figure 7c has a maximum. The α-helix structure contained eight hydrogen bonds. Of these, six followed the standard α-helix pattern, connecting the peptide C—O of residue i with the peptide N—H of residue $i + 4$; the additional two hydrogen bonds involved the terminal blocking groups. The 3_{10}-helix structure contained nine hydrogen bonds; seven were $i - i + 3$ bonds typical of the 3_{10}-helix, and two involved terminal blocking groups. The 3_{10}-helix hydrogen bond geometry was significantly distorted from the linear arrangement found in the α-helix. In the intermediate structure the hydrogen bonds at the helix ends—those involving C—O groups of residues 1, 6, and 7, as well as the blocking groups—were similar to the 3_{10}-helix state, while the remaining peptide C—O groups formed bifurcated hydrogen bonds, involving both $i - i + 3$ and $i - i + 4$ contacts.

AIB10 FREE ENERGY GRADIENT MAP

The reduced, two-dimensional free energy gradient map describing concerted transitions of ϕ/ψ dihedrals in the right-handed helical region of the AIB10 peptide (CH_3CO-$(Aib)_{10}$-$CONH_2$) is presented in Figure 7. As in the ALA10 case, we find two regions of low gradient corresponding to the α-helix and 3_{10}-helix structures. The α-helix region is a small trough near (−57°, −50°) running diagonally across the lower right-hand corner of the map. This region is shifted by about 6° toward

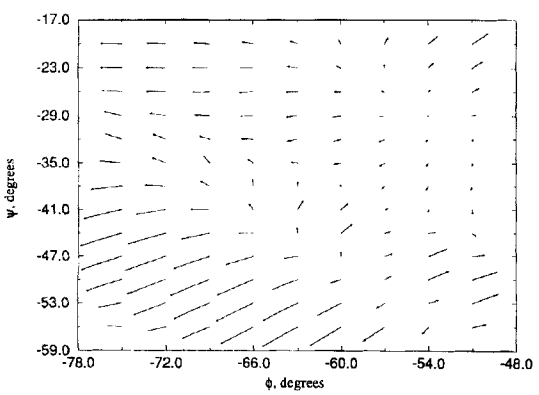

FIGURE 7. The reduced two-dimensional free energy gradient map of the AIB10 peptide. Scale is identical to that in Figure 4.

higher ϕ and lower ψ relative to ALA10. The 3_{10}-helix region is also shifted to higher ϕ relative to ALA10, and lies in the vicinity of (−54°, −29°). As in the ALA10 case, we identify the minima of the free energy surface, evaluate free energy profiles, and perform free energy decompositions.

From the reduced two-dimensional map shown in Figure 7 two free energy minima can be identified, in the vicinity of (−57°, −50°) in the α-helix region, and close to (−54°, 29°) in the 3_{10}-helix region. The reduced map indicates that α-helix and 3_{10}-helix conformations in those areas are stable with respect to small displacements in which all ϕ and/or all ψ dihedrals are changed in concert, $\Delta\phi_i = \Delta\phi$, $\Delta\psi_i = \Delta\psi$, $i = 1,\ldots,9$. Analysis of the full 18-dimensional free energy gradient shows that all gradient components change sign from negative to positive as we pass through the (−57°, −50°) and (−54°, −29°) points, indicating that minima of the complete 18-dimensional free energy hypersurface exist close to the minima of the reduced surface. Thus, both the α-helix and the 3_{10}-helix are stable molecular states of the AIB10 peptide in vacuum.

Free energy profiles obtained from the AIB10 free energy gradient map are presented in Figures 5 and 6. The profile obtained by integrating the free energy gradient along the α-helix trough (Fig. 5) indicates that the conformation of lowest free energy is the α-helical structure close to (−55.5°, −51.5°), in good agreement with the average conformation of (−57°, −50°) found in Aib homopeptide crystals.[62] For AIB10 anticorrelated changes of ϕ, ψ by ±4.5° lead to free energy changes of about 1.2 kcal/mol, about twice as much as found for ALA10. The α-methyl substitution, which decreases the conformational flexibility of the system, thus makes the α-helix more rigid, in accord with chemical intuition. Figure 6 shows the free energy profile along the $\psi = -54°$ line, which connects the α-helix and 3_{10}-helix regions. The structure with lowest free energy in the 3_{10}-helix region is close to (−54°, −29°). This is in excellent agreement with average crystal conformations of (−54°, −28°) found in Aib homopeptide 3_{10}-helices.[63]

Integrating the AIB10 free energy gradient from (−55.5°, −51.5°) to (−54°, −53°) and then along $\phi = -54°$ to (−54°, −29°), we find that within our model the α-helix is more stable than the 3_{10}-helix by 3.9 ± 0.1 kcal/mol for the AIB10 peptide in vacuum. In Figure 6, the barrier for the concerted transition between the two AIB10 helical states is about 3.1 ± 0.1 kcal/mol above the 3_{10}-helix. The

smaller α-helix/3_{10}-helix free energy difference in AIB10 relative to ALA10 is in qualitative agreement with experimental observations of both α-helical and 3_{10}-helical conformations in Aib-containing peptides (see Ref. 33 for a discussion). In an umbrella sampling simulation, Marshall et al. found the α-helix more stable than the 3_{10}-helix by 1.1 kcal/mol in the Aib nonapeptide; the crossing of the 0.8 kcal/mol 3_{10}-helix → α-helix barrier occurred sequentially.[33] Our free energy difference between the two helical states of AIB10 is consistent with Marshall's result, since it is estimated that the stability of the α-helix in longer peptides increases relative to the 3_{10}-helix by 2–3 kcal/mol per residue.[33] The lower barrier between the 3_{10}-helix and α-helix obtained by Marshall et al. indicates that the sequential transition should be favored over the concerted one in Aib homopeptides.

The hydrogen bonding patterns in energy-minimized AIB10 structures representative of the α-helix (−57°, −54°), 3_{10}-helix (−54°, −29°), and the intermediate form (−54°, −44°) are similar as those obtained for ALA10. The α-helix is characterized by $i - i + 4$ hydrogen bonds, the 3_{10}-helix by $i - i + 3$ bonds, and the intermediate by $i - i + 3$ bonds at the helix ends and bifurcated bonds in the central part of the helix.

Comparison of the ALA10 and AIB10 results allows to draw some general conclusions of the influence of the α-methyl substitution on peptide conformations. The main effect is the stabilization of the 3_{10}-helix structure. The 3_{10}-helix is unstable for ALA10 and becomes a stable state for AIB10, the α-helix/3_{10}-helix free energy difference decreases from about 14 to about 4 kcal/mol. Further changes include a shift of the α-helix and 3_{10}-helix states toward higher ϕ in AIB10 relative to ALA10 and an increased rigidity of the α-helix with respect to anticorrelated changes of ϕ and ψ in AIB10 compared to ALA10.

Our ALA10 and AIB10 simulations demonstrate the feasibility evaluating multidimensional free energy gradients for large molecular systems. The potential usefulness of the CFTI method in exploring multidimensional free energy surfaces goes well beyond the relatively simple applications presented here. Taking a page from the well-developed methods of studying potential energy surfaces (e.g., Ref. 64), analogous methods of free energy surface analysis may be developed—free energy minimization, transition state optimization, and determination of minimum free energy paths between two states. Approaches using these methods will enable precise determination of free energy minima, transition states and pathways for conformational change. Although the ALA10 and AIB10 test cases were simulated in vacuum, calculations of molecules in solution may readily be undertaken within the same formalism, requiring only more computer time. The CFTI method provides an effective tool for calculating the gradient of the free energy with respect to all selected conformational coordinates of interest, not just along a known or guessed reaction coordinate. Thus, working within the multidimensional CFTI approach guarantees that the optimized minima and transition states correspond to the lowest possible free energy with respect to changes in the selected coordinate set.

Conclusions

A novel approach to conformational free energy simulations—the CFTI method—is presented. The algorithm allows treatment of both one-dimensional and multidimensional cases, enabling the description of multidimensional free energy surfaces. Analysis of the thermodynamic integration approach leads to a computationally efficient method of evaluating free energy gradients with respect an arbitrary number of conformational coordinates. Methods of evaluating higher derivatives of the free energy are also proposed. The CFTI approach has a number of interesting properties: (i) CFTI uses the local energy gradient, which is easily available in standard molecular simulations; (ii) evaluations of multidimensional free energy gradients carry very little overhead compared to evaluating the derivative with respect to a single reaction coordinate, and multidimensional CFTI calculations can be performed at practically the same cost as standard molecular dynamics simulations; (iii) the CFTI approach allows the calculation of second and higher derivatives of the conformational free energy, which can enhance the rate of exploration of the free energy surface; and (iv) CFTI free energy differences can be decomposed into contributions from different energy terms (internal deformation, van der Waals, electrostatic) and parts of the system (solute, solvent, individual chemical groups), providing insight into the microscopic mechanisms of various underlying molecular conformational equilibria.

A related algorithm for energy minimization of molecular systems under holonomic constraints is also developed and described, which allows for generation of adiabatic energy maps in the constrained coordinates.

The general CFTI algorithm is applied here to study the conformational equilibrium of *n*-butane in the gas and liquid phases, and to characterize the multidimensional conformational free energy surfaces of the decaalanine (ALA10) and deca-α-methylalanine (AIB10) peptides. For *n*-butane, the CFTI results are compared with those from the thermodynamic perturbation (TP) method and direct molecular dynamics simulations using the same butane model, as well as available experimental data and previous computational results on different models. The three simulation methods employed here give essentially identical results for the *n*-butane conformational free energy profile in the gas phase and liquid. Our results thus indicate that, for the one-dimensional example of *n*-butane, the CFTI approach yields overall results of similar quality to other established methods.

The *gauche–trans* equilibrium constants calculated using the CHARMM (Version 22) explicit hydrogen butane model[36] are 0.3–0.4, which puts them in or below the lower range of experimentally observed and previously computed values. This is consistent with an increase of the adiabatic energy difference between the *gauche* and *trans* conformers in the new CHARMM (Version 22) parameter set relative to those used previously. Within the established error range, our simulations predict that there is essentially no shift in the conformational equilibrium of *n*-butane between the gas and liquid. This is in agreement with some simulation results[29,30,55,56] and in contrast to other studies which found a preference for the *gauche* form in liquid butane and different butane solutions.[16,20,37,43–45,57] Since this lack of equilibrium shift is present in our CFTI, TP, and direct simulation results, it is clearly a property of the molecular model used, which differs in details of description of internal deformations and nonbonded interactions from those used in other reported butane studies.

The uncertainties in entropic components of the free energy profile are much larger than those of the free energies, limiting the usefulness of the energy–entropy decomposition, as has been found before.[20] In contrast, decompositions of the free energy into components from butane internal deformations, and intramolecular and intermolecular nonbonded interactions gave results of comparable precision to the total free energies. This decomposition analysis was used to detect and characterize conformational crowding in the *cis* transition state of *n*-butane.

For the two decapeptides studied, the full 18-dimensional free energy gradient was used to generate reduced, two-dimensional free energy gradient maps for concerted changes in the ϕ and ψ dihedrals. For decaalanine, the α-helix was found to be the only stable molecular state in the studied region. The α-helix structures occupied a relatively long trough, with soft deformations for anticorrelated changes of ϕ and ψ and hard deformations for correlated changes. The α-helix was found to be more stable than the 3_{10}-helix by about 13.7 kcal/mol, in good agreement with previous calculations using a different model.[32] A number of differences between ALA10 and AIB10 conformational free energy surfaces was found, attributable to the effects of α-methyl substitution. The most important difference was that both the α-helix and the 3_{10}-helix were stable molecular states for AIB10, in qualitative agreement with observations of both types of helical conformations in Aib-containing peptides. The α-helix trough in AIB10 was steeper and shifted toward higher ϕ values compared to ALA10, in accord with crystallographic results. Finally, the free energy difference between the α-helix and 3_{10}-helix states was only 3.9 kcal/mol in AIB10, while the barrier for the concerted 3_{10}-helix → α-helix transition was 3.1 kcal/mol. The free energy difference between the α-helix and 3_{10}-helix conformers of AIB10 found here is consistent with umbrella sampling simulation results of the sequential 3_{10}-helix → α-helix transition in the Aib nonapeptide of Marshall et al.[33] The lower transition barrier obtained in Marshall et al.'s work indicates that the sequential transition between the two helical states should be favored for the Aib homopeptides of length close to ten residues. In both ALA10 and AIB10, intermediates on the pathway for the concerted α-helix → 3_{10}-helix transition had 3_{10}-helix type hydrogen bonding at the helix ends and bifurcated hydrogen bonds in the central part.

In summary, based on the presented simulations, we can conclude that the CFTI method systematically gives numerical results of comparable quality to other established conformational free energy simulation algorithms. With its algorithmic simplicity, computational efficiency, possibility of

decomposition analysis of free energy differences, and generalizations to many dimensions and calculations of higher free energy derivatives, the CFTI approach emerges as a highly useful tool for present and future investigations of conformational equilibria of flexible molecules in condensed phases. Especially interesting is the application of the CFTI approach to characterize multidimensional conformational free energy surfaces. The potential usefulness of this method goes well beyond the relatively simple applications described here. Extensions of the current studies toward simulations including solvent and employing more refined free energy surface exploration algorithms are underway.

Acknowledgments

The author thanks Professor Martin Karplus for making available CHARMM (Version 22) parameters prior to publication. This work was supported in part by the Kansas Institute for Theoretical and Computational Science.

Appendix A: Derivation of Basic TI Formula

Starting with the joint probability distribution $\rho(\xi_1', \xi_2', \ldots, \xi_m')$ given in eq. (14), we can define the "slice" configuration integral $Z(\xi_1', \xi_2', \ldots, \xi_m')$ in analogy in eq. (5):

$$\rho(\xi_1', \xi_2', \ldots, \xi_m') = Z(\xi_1', \xi_2', \ldots, \xi_m')/Z \quad (19)$$

The multidimensional conformational free energy surface is then given by:

$$A_c(\xi_1', \xi_2', \ldots, \xi_m') = -\beta^{-1} \ln Z(\xi_1', \xi_2', \ldots, \xi_m') \quad (20)$$

and its gradient is:

$$\frac{\partial A_c(\xi_1', \xi_2', \ldots, \xi_m')}{\partial \xi_k'} = -\beta^{-1} \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \times \frac{\partial Z(\xi_1', \xi_2', \ldots, \xi_m')}{\partial \xi_k'} \quad (21)$$

To calculate the derivative of the "slice" configuration integral $Z(\xi_1', \xi_2', \ldots, \xi_m')$ with respect to ξ_k' we will derive a general formula for any function $X(q)$:

$$\frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \frac{\partial}{\partial \xi_k'} [Z(\xi_1', \xi_2', \ldots, \xi_m') \langle X(q) \rangle_{\xi'}] = \left\langle \frac{\partial}{\partial \xi_k'} [X(q) e^{-\beta U(q)}] e^{\beta U(q)} \right\rangle_{\xi'} + \left\langle X(q) \frac{\partial \ln |J|}{\partial \xi_k'} \right\rangle_{\xi'} \quad (22)$$

where $\langle \cdots \rangle_{\xi'}$ denotes the canonical average over the subset of conformations with fixed values of $\xi_i = \xi_i'$, $i = 1, \ldots, m$ and J is the Jacobian of the transformation from the coordinates q to the generalized coordinates u of which ξ_i, $i = 1, \ldots, m$ are a subset, $u = (\eta; \xi_1, \ldots, \xi_m)$. To prove eq. (22) we use the definition of the derivative on the left-hand side:

$$\begin{aligned}
&\frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \frac{\partial}{\partial \xi_k'} [Z(\xi_1', \xi_2', \ldots, \xi_m') \langle X(q) \rangle_{\xi'}] \\
&= \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \frac{\partial}{\partial \xi_k'} \int dq\, X(q) \times e^{-\beta U(q)} \prod_{i=1}^{m} \delta(\xi_i(q) - \xi_i') \\
&= \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \frac{\partial}{\partial \xi_k'} \int du |J(u)| X(u) \times e^{-\beta U(u)} \prod_{i=1}^{m} \delta(\xi_i - \xi_i') \\
&= \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \lim_{\Delta\xi \to 0} \frac{1}{\Delta\xi} \times \int du |J(u)| X(u) e^{-\beta U(u)} \\
&\quad \times [\delta(\xi_k - \xi_k' - \Delta\xi) - \delta(\xi_k - \xi_k')] \times \prod_{i \neq k} \delta(\xi_i - \xi_i') \\
&= \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \lim_{\Delta\xi \to 0} \frac{1}{\Delta\xi} \int du \times [|J(u^*)| X(u^*) e^{-\beta U(u^*)} \\
&\quad - |J(u)| X(u) e^{-\beta U(u)}] \prod_{i=1}^{m} \delta(\xi_i - \xi_i') \\
&= \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \int du \frac{\partial}{\partial \xi_k'} \times [|J(u)| X(u) e^{-\beta U(u)}] \prod_{i=1}^{m} \delta(\xi_i - \xi_i')
\end{aligned}$$

$$= \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \int dq \frac{1}{|J|} \frac{\partial}{\partial \xi_k'} \times [|J(q)| X(q) e^{-\beta U(q)}] \prod_{i=1}^{m} \delta(\xi_i(q) - \xi_i')$$

$$= \frac{1}{Z(\xi_1', \xi_2', \ldots, \xi_m')} \int dq \times \left[\frac{\partial}{\partial \xi_k'} (X(q) e^{-\beta U(q)}) + X(q) \frac{\partial \ln|J(q)|}{\partial \xi_k'} e^{-\beta U(q)} \right] \times \prod_{i=1}^{m} \delta(\xi_i(q) - \xi_i')$$

$$= \left\langle \frac{\partial}{\partial \xi_k'} [X(q) e^{-\beta U(q)}] e^{\beta U(q)} \right\rangle_{\xi'} + \left\langle X(q) \frac{\partial \ln|J|}{\partial \xi_k'} \right\rangle_{\xi'} \quad (23)$$

in the transformations above $|J(u)| X(u) e^{-\beta U(u)} \delta(\xi_k - \xi_k' - \Delta\xi)$ was replaced by the equivalent expression $|J(u^*) X(u^*) e^{-\beta U(u^*)} \delta(\xi_k - \xi_k')$; u^* is the perturbed configuration u, corresponding to changing $\xi_k \rightarrow \xi_k + \Delta\xi$ and leaving all the other degrees of freedom unchanged. The derivative of any function of q with respect to ξ_k such as that appearing in the last two lines of eq. (23) may be evaluated by the chain rule, e.g., for $X(q)$:

$$\frac{\partial X(q)}{\partial \xi_k} = \sum_{i=1}^{3N} \frac{\partial X(q)}{\partial q_i} \frac{\partial q_i}{\partial \xi_k} \quad (24)$$

Appendix B: Implementation of Energy Minimization with Holonomic Constraints

A set of holonomic constraints may be defined as:

$$\Xi_i(q) - \xi_i = 0, \qquad i = 1, \ldots, m$$

where the $\Xi_i(q)$ are some scalar functions of the Cartesian coordinates and ξ_i are the values at which they have been fixed. To enable efficient energy minimization of a molecular system under the above constraints, it is necessary to perform two tasks: (a) reset the coordinates to satisfy constraints after every minimization step; and (b) make the search direction of the minimization algorithm orthogonal to the gradient of the constraints. Implementing (a) only leads to slow convergence as the degrees of freedom orthogonal to constraints relax and major components of minimization steps are rejected. Additional problems appear with satisfying standard convergence criteria of minimization algorithms, since the gradient does not converge to zero. On the other hand, implementing (b) only leads to violation of constraints due to the finite size of the minimization steps.

The new method of potential energy minimization with holonomic constraints developed by the author uses both steps (a) and (b), analogously, to the available option of energy minimization with SHAKE constraints.[35] Within the algorithm each energy minimization step consists of the following stages:

1. Reset coordinates to satisfy constraints using the algorithm of Tobias and Brooks.[40]
2. Evaluate the potential energy U and its Cartesian gradient $\partial U/\partial q_j$ at the current position.
3. Orthogonalize energy gradient $\partial U/\partial q_j$ to the gradients of the constraints. This is done by subtracting out the energy gradient components along the constraint coordinates. The subtraction is repeated iteratively until the remaining gradient is orthogonal to all constraints within a desired accuracy. This procedure is analogous to the SHAKE algorithm for coordinate constraints.[38]
4. If a conjugate-gradient-type method is used, orthogonalize the search directions to the constraints analogously as described above.
5. If minimization has not converged, perform an energy minimization step and go to stage 1; otherwise stop.

The algorithm works with all methods using only energy and gradient information available in CHARMM—steepest descent, conjugate gradient, Powell, and adopted-basis Newton–Raphson.[35] For the difficult test case of the decapeptide in vacuum with all backbone dihedrals fixed, in which there is significant overlap between the constraints, the orthogonalization in stages 3–4 converges in about 50 iterations. The operations involved in the orthogonalization are mainly vector algebra—calculation of scalar products and norms and vector addition and are of similar complexity as SHAKE constraints.

The new constrained energy minimization algorithm was implemented by the author through extending and modifying the standard CHARMM (Version 22) program package.

Appendix C: Implementation of CFTI Method

The new thermodynamic integration conformational free energy (CFTI) algorithm consists of the following steps.

1. *Trajectory generation.* A molecular dynamics trajectory with selected internal coordinates to remain at fixed values is generated using the holonomic constraint method of Tobias and Brooks.[40]

2. *Gradient evaluation.* The derivative of the potential energy with respect to a conformational coordinate, ξ_k, was calculated along the trajectory according to eq. (24):

$$\frac{\partial U}{\partial \xi_k} = \sum_{i=1}^{3N} \frac{\partial U}{\partial x_i} \frac{\partial x_i}{\partial \xi_k} \tag{25}$$

where the x_i are the Cartesian coordinates of the system. The terms $\partial U/\partial x_i$ are just the components of the potential energy gradient, and are available at every molecular dynamics integration time step. The quantities that need to be computed specifically for the CFTI method are the coefficients $\partial x_i/\partial \xi_k$, $i = 1,\ldots,3N$, showing how the Cartesian coordinates of atoms change along the conformational coordinate ξ_k.

For the change of position of atom j due to an infinitesimal right-handed rotation by angle $d\phi$ around an axis with unit vector $\vec{n}$ we use the standard formula[65]:

$$d\vec{r}_j = \vec{n} \times \left(\vec{r}_0 - \vec{r}_j\right) \cdot d\phi \tag{26}$$

where $\vec{r}_j = (x_j, y_j, z_j)$ is the position of an atom j and $\vec{r}_0 = (x_0, y_0, z_0)$ is the position of a point lying on the rotation axis. For this type of coordinate, relatively few atoms will have nonzero coefficients $\partial x_i/\partial \phi$. For example, for rotation around the central C—C bond of butane in solution, only the methylene hydrogen and methyl group positions vary with ϕ, while the coordinates of the central bond carbons and the solvent are not changed.

3. *Storage of crucial data.* The calculated $\partial U/\partial \xi_k$, $k = 1,\ldots,m$ values, together with the values of the potential and kinetic energies are saved in an auxiliary disk file at a given frequency during the trajectory generation phase. For one-dimensional profiles, additional entries in the same file contain information about the energy of the system perturbed by small increments $\xi \pm \Delta\xi$ of the conformational coordinate. This enables evaluation of free energy changes using both thermodynamic integration (TI) and thermodynamic perturbation (TP) methods in the postprocessing stage. For multidimensional surfaces the TP method was not implemented, as the additional energy evaluations it requires make it inefficient to use in the general case.

Additionally, system coordinates are saved along the trajectory for use in the CFTI decomposition analysis.

4. *Postprocessing.* At this stage, the data stored on disk is used to obtain free energy profiles and their various decompositions. The trajectory averages of $\partial U/\partial \xi_k$, $k = 1,\ldots,m$ are calculated, yielding the free energy gradient with respect to the selected set of conformational coordinates [eqs. (10) and (15)] at the fixed values of ξ_k for which the simulation was performed. Trajectory averages of $\partial U/\partial \xi$, the potential energy U, and $U \cdot (\partial U/\partial \xi)$ are used to evaluate the derivative of the conformational entropy with respect to ξ [eq. (13)]. The derivatives of the free energy and entropy with respect to the conformational coordinate ξ are then numerically integrated over ξ starting from a chosen reference conformation ξ_0, to yield the corresponding profiles $\Delta A_c(\xi)$ and $-T\Delta S_c(\xi)$. Finally, the energy profile $\Delta U(\xi)$ is calculated using eq. (12).

The alternative method of energy–entropy decomposition using the value of $\langle U \rangle_\xi$ calculated from the trajectory and setting $-TS_c(\xi) = A_c(\xi) - \langle U \rangle_\xi$ was abandoned, since it leads to statistical errors much larger than those from the implemented protocol. In calculating $\Delta U(\xi)$ from the derivative $\partial \langle U \rangle_\xi / \partial \xi$, we are eliminating significant uncertainties in the calculated quantity. The reason for this is the same as that for the high precision of free energy values in general—we are calculating only differences of properties between states, not absolute values, and restricting our attention only to that part of the potential energy which varies with the conformational coordinate, eliminating solvent–solvent interactions from considerations.

Further decompositions of the free energy profile are performed by reading in the stored trajectory coordinates, recalculating energies and gradients with only the selected energy terms activated (e.g., only internal deformation terms or only in-

tramolecular nonbonded interactions), and evaluating the trajectory average contribution to $\partial U/\partial \xi$.

In the one-dimensional case, the data stored in the auxiliary file in step 3 make possible the use of the TP protocol Tobias and Brooks[20] to calculate free energy; energy and entropy differences between the fixed value, ξ, for which the simulation was performed; and the perturbed configurations $\xi \pm \Delta\xi$. The free energy profile is then obtained by adding up the differences starting from a chosen reference conformation ξ_0. Energy and entropy profiles are also calculated from the auxiliary file using the numerical temperature derivative of the partition function.[20]

In the TP and TI protocols, the statistical errors of the trajectory averages involved in the individual free energy, entropy and energy differences (TP), or derivatives (TI) are calculated by dividing the trajectory points into a number of contiguous subsets (blocks) and evaluating the standard deviation of the mean of the block subaverages. The statistical errors of the free energy, entropy, and energy profiles were calculated using standard error propagation methods.

The new CFTI algorithm was implemented by the author through extending and modifying the standard CHARMM (Version 22) program package.

Appendix D: Error Analysis

In this section, the magnitudes and sources of errors in the calculated quantities are briefly analyzed. For more details on calculation of statistical errors see Appendix C.

The fundamental quantities obtained from the CFTI simulations are the derivatives of the conformational energy with respect to conformational coordinates, obtained as trajectory averages of corresponding potential energy derivatives. For the decapeptide 40-ps vacuum trajectories the typical statistical errors of individual derivatives $\partial A_c/\partial \phi_i$, $\partial A_c/\partial \psi_i$, $i = 1,\ldots,9$, were about 0.1–0.2 kcal/(mol rad), leading to errors of the aggregate derivatives $\partial A_c/\partial \phi$, $\partial A_c/\partial \psi$ of about 0.5 kcal/(mol rad), and translating to uncertainties of integrated free energy differences across the full range of the map of 0.1 kcal/mol or less. The precision of the nonbonded components was similar to that of the total free energy, while the internal strain contributions exhibited smaller fluctuations.

For *n*-butane the typical statistical errors for averages over a 40-ps trajectory were 0.02 and 0.05 kcal/(mol rad) in the gas phase and liquid, respectively. After integration of the derivative this translates into statistical errors of the free energy difference between the *gauche* and *trans* conformers of 0.01 (0.02) kcal/(mol rad) in the gas (liquid). Understandably, the errors for the *syn* conformer were larger in magnitude. In the TP method, for 5° changes in ϕ, the typical errors in ΔA_c were 0.003 (0.007) kcal/mol in the gas (liquid) simulations, leading to accumulated errors of 0.02 (0.04) for the *gauche* free energy relative to *trans*. In the direct MD simulations, the typical errors of the symmetrized distributions were about 0.1% closer to the highly populated *trans* state, about 3% in the *gauche* region, and above 50% close to the *syn* maximum. The statistical uncertainties of the free energy profiles are quite similar for the three methods except in the vicinity of the *syn* barrier, where the free energy simulation methods exhibit much better sampling than the direct MD approach.

The second quantity of interest obtained from the CFTI simulations is the derivative $-T\partial S_c/\partial \phi$ calculated according to eq. (13). Its typical statistical errors were 0.06 (2.0) kcal/(mol rad) for *n*-butane in the gas phase (liquid). After integration of the derivative this translates into statistical errors of the entropic component difference between the *gauche* and *trans* conformers of 0.04 (1.00) kcal/mol, with still larger errors for the *syn* conformer. For states differing in ϕ by 5° the typical statistical error of the TP entropy change was 0.1 (0.9) kcal/mol in the gas (liquid) simulations, leading to an accumulated error of 0.4 (3.6) kcal/mol for the *gauche* conformational entropy relative to *trans*. In both TI and TP methods, the entropies have much larger statistical errors than the free energies. Also, the TP entropy uncertainties are several times larger than those from TI. The entropy formulae for both methods [eq. (13) and Ref. 20] involve evaluation of the total potential energy of the simulated system which includes both terms directly varying with the solute conformation (solute–solute and solute–solvent) and the solvent–solvent energy, which does not change upon infinitesimal perturbation of the conformational coordinate. It appears that the approach used in the CFTI method is able to better eliminate the influence of the large solvent–solvent term and give more precise entropy components (see also Appendix A). The TI and TP conformational ener-

gies $\Delta U(\phi)$ have statistical errors essentially identical to those of the entropies $-T\Delta S(\phi)$.

In the TI approach, the free energy profile may also be decomposed into contributions from the different parts of the system and potential energy terms: butane internal deformations and nonbonded interactions. In the gas phase, typical errors of the internal deformation and intramolecular nonbonded terms were 0.02–0.03 kcal/(mol rad), essentially identical to those of the total free energy. This translates into uncertainties of about 0.01 kcal/mol to the *gauche–trans* free energy difference. In the liquid, the statistical errors of the internal deformation and intramolecular and intermolecular interaction components were typically 0.02, 0.06, and 0.25 kcal/(mol rad), respectively, leading to uncertainties of the *gauche–trans* free energy differences in these components of 0.01, 0.03, and 0.12 kcal/mol. The intermolecular interaction term, which is the most interesting one, has the largest uncertainty. Because of its small overall magnitude, the intermolecular contribution cannot be distinguished from zero within the errors. The internal deformation and intramolecular terms are within errors essentially the same as the corresponding gas phase results.

The statistical errors of the *trans–gauche* equilibrium constants from the TI and TP free energy simulations are about 0.005 (or 0.3% *trans* fraction) in both the gas and liquid. The statistical errors of the direct MD equilibrium constants calculated from the asymmetry of the dihedral distribution were of similar magnitude; errors evaluated using the block average approach, as half of the difference between the results from the first and second halves of the trajectories, were significantly greater: 0.035 (2% *trans* fraction) in the gas and 0.015 (0.8% *trans* fraction) in the liquid. Thus, it appears that the asymmetry method tends to underestimate the statistical errors of the direct simulations. Using the block average method and error propagation $\delta A = kT \cdot \delta P/P$, the direct MD free energy errors reported in Table I should be multiplied by a factor of 4 for the gas phase and 2 for the liquid.

Besides statistical errors, the free energy simulation results also exhibit systematic errors. The main source of systematic error are temperature fluctuations. In the *n*-butane gas phase free energy simulations, the average temperature of the 760-ps production phase was 275 ± 50 K, i.e., the temperature fluctuations over the whole trajectory were 50 K. However, the 19 individual fragments of 40-ps length, used to evaluate TI free energy derivatives and TP free energy changes, did not all correspond to a temperature of 275 K. Actually, the standard deviation of the sample of average temperatures from the 19 fragments was 16 K. In the liquid the average temperature over the 760-ps production phase trajectory was 272 ± 6 K; the standard deviation of the sample of 19 fragments of 40-ps length was 2 K. Thus, the free energy results from the individual windows correspond to different effective temperatures. To estimate the influence of this effect on the accuracy of the results we resort to an approximate procedure. Using the temperature-independent adiabatic profile $\Delta U_m(\phi)$ we can obtain an approximate variation of the *trans–gauche* equilibrium constant K_{tg} with temperature: increasing temperature by 1 K raises K_{tg} by about 0.002, and decreases the *trans* fraction by about 0.1%. Using the information we can estimate that temperature fluctuations should introduce an error of about 0.03 to the equilibrium constant in the gas phase and increase the errors of the TI and TP free energies in Table I by a factor of 2 to 3. In the liquid the effect of temperature fluctuations is negligible.

Taking into account the enhanced error estimates presented above the free energy profiles $\Delta A_c(\phi)$ from the three methods, TI, TP, and MD, and the two phases, gas and liquid, are all essentially identical. Within these error estimates we find equilibrium constants of 0.31 ± 0.04, 0.30 ± 0.03, and 0.34 ± 0.03 from the direct MD, TI, and TP methods, respectively, in the gas phase. The corresponding liquid results are 0.37 ± 0.02, 0.28 ± 0.01, and 0.31 ± 0.01. The gas and liquid equilibrium constants given by each method are thus identical within the errors; our results predict essentially no equilibrium shift between the gas and liquid. Other systematic errors in the free energy results are the effect of finite size of the dihedral perturbation in TP, and the numerical integration of the derivatives in TI. Since corresponding TI and TP results agree with each other within the established error margins, it appears that these effects are not significant.

Two interesting observations may be made based on this analysis. First, there is significant difficulty in obtaining precise values of equilibrium constants for flexible *n*-butane in the gas phase using all three methods applied here. The reason for this lies in the large relative fluctuations of the kinetic energy (and thus temperature) for this small system. Second, in the liquid simulations, the direct MD approach yielded an equilibrium constant greater than results of both em-

ployed free energy methods; the differences are greater than the errors established here. Since the free energy profiles from the three methods agree within their error ranges, it appears that the equilibrium constants are quite sensitive to detailed shapes of the profiles.

References

1. M. Mezei and D. L. Beveridge, *Ann. NY Acad. Sci.*, **482**, 1–23 (1986).
2. D. L. Beveridge and F. M. DiCapua, *Annu. Rev. Biophys. Chem.*, **18**, 431–492 (1989).
3. T. P. Straatsma and J. A. McCammon, *Annu. Rev. Phys. Chem.*, **43**, 407–435 (1992).
4. J. Gao, K. Kuczera, B. Tidor, and M. Karplus, *Science*, **244**, 1069–1072 (1989).
5. W. L. Jorgensen and C. Ravimohan, *J. Am. Chem. Soc.*, **83**, 3050–3054 (1985).
6. P. A. Bash, U. C. Singh, R. Langridge, and P. A. Kollman, *Science*, **236**, 564–568 (1987).
7. T. P. Lybrand, I. Ghosh, and J. A. McCammon, *J. Am. Chem. Soc.*, **107**, 7793–7794 (1985).
8. C. L. Brooks III, *J. Phys. Chem.*, **90**, 6680–6684 (1986).
9. T. P. Lybrand, J. A. McCammon, and G. Wipff, *Proc. Natl. Acad. Sci. USA*, **83**, 833–835 (1986).
10. P. A. Bash, M. J. Field, and M. Karplus, *J. Am. Chem. Soc.*, **109**, 8092–8094 (1987).
11. M. J. Field, P. A. Bash, and M. Karplus, *J. Comput. Chem.*, **11**, 700–733 (1990).
12. C. F. Wong and J. A. McCammon, *J. Am. Chem. Soc.*, **108**, 3830–3832 (1986).
13. V. Daggett, F. Brown, and P. A. Kollman, *J. Am. Chem. Soc.*, **111**, 8247–8256 (1989).
14. B. Tidor and M. Karplus, *Biochemistry*, **30**, 3217–3228 (1991).
15. K. Kuczera, J. Gao, B. Tidor, and M. Karplus, *Proc. Natl. Acad. Sci. USA*, **87**, 8481–8485 (1990).
16. H. Hayashi, H. Tanaka, and K. Nakanishi, *Molec. Simul.*, **9**, 401–415 (1993).
17. J. P. Valleau and G. M. Torrie, In B. J. Berne, Ed., *Statistical Mechanics, Part A, Equilibrium Techniques*, Plenum, New York, 1977.
18. W. F. van Gunsteren and H. J. C. Berendsen, *Angew. Chem. Int. Ed. Engl.*, **29**, 992–1023 (1990).
19. W. L. Jorgensen, *J. Phys. Chem.*, **87**, 5304–5314 (1983).
20. D. J. Tobias and C. L. Brooks III, *J. Chem. Phys.*, **92**, 2582–2592 (1990).
21. T. P. Straatsma and J. A. McCammon, Multiconfiguration thermodynamic integration, *J. Chem. Phys.*, **95**, 1175–1188 (1991).
22. D. E. Smith and A. D. J. Haymet, *J. Chem. Phys.*, **98**, 6445–6454 (1993).
23. L. R. Pratt, C. S. Hsu, and D. Chandler, *J. Chem. Phys.*, **68**, 4202–4212 (1978).
24. E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, *Chem. Phys. Lett.*, **156**, 472–477 (1989).
25. D. Chandler, *Faraday Disc. Chem. Soc.*, **66**, 184–190 (1978).
26. C. L. Brooks III and D. A. Case, *Chem. Rev.*, **93**, 2487–2502 (1993).
27. W. Feller, *Introduction to Probability Theory and Its Applications*. 3rd ed., Wiley, New York, 1967.
28. D. A. McQuarrie, *Statistical Mechanics*, Harper and Row, New York, 1976.
29. D. Brown and J. H. Clarke, *J. Chem. Phys.*, **92**, 3062–3073 (1990).
30. W. L. Jorgensen, *J. Am. Chem. Soc.*, **103**, 4721–4726 (1981).
31. J. Hermans, *Curr. Opin. Struct. Biol.*, **3**, 270–276 (1993).
32. J. Tirado-Rives, D. S. Maxwell, and W. L. Jorgensen, *J. Am. Chem. Soc.*, **115**, 11590–11593 (1993).
33. S. E. Huston and G. E. Marshall, *Biopolymers*, **34**, 75–90 (1994).
34. M. L. Smythe, S. E. Huston, and G. E. Marshall, *J. Am. Chem. Soc.*, **117**, 5445–5452 (1995).
35. B. R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.*, **4**, 187–217 (1983).
36. A. D. MacKerrell, Jr., J. Wiórkiewicz-Kuczera, and M. Karplus, *J. Am. Chem. Soc.*, **117**, 11946–11975 (1995).
37. G. Kaminski, E. M. Duffy, T. Matsui, and W. L. Jorgensen, *J. Phys. Chem.*, **98**, 13077–13082 (1994).
38. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comp. Phys.*, **23**, 327–341 (1977).
39. J. C. Smith and M. Karplus, *J. Am. Chem. Soc.*, **114**, 801–812 (1992).
40. D. J. Tobias and C. L. Brooks III, *J. Chem. Phys.*, **89**, 5115–5127 (1988).
41. M. P. Allen and D. J. Tildesley, *Computer Simulations of Liquids*. Oxford, London, 1987.
42. E. Barth, K. Kuczera, B. Leimkuhler, and R. D. Skeel, *J. Comp. Chem.*, **16**, 1192–1209 (1995).
43. J.-P. Ryckaert and A. Bellemans, *Discuss. Faraday Soc.*, **66**, 95–106 (1978).
44. R. Edberg, D. J. Evans, and G. P. Moriss, *J. Chem. Phys.*, **84**, 6933–6939 (1986).
45. P. A. Wielopolski and E. R. Smith, *J. Chem. Phys.*, **84**, 6940–6942 (1986).
46. J. R. Durig, A. Wang, W. Beshir, and T. S. Little, *J. Raman Spectrosc.*, **22**, 683–704 (1991).
47. S. Kint, J. Scherer, and R. G. Snyder, *J. Chem. Phys.*, **73**, 2599 (1980).
48. L. Rosenthal, J. F. Robolt, and J. Hummel, *J. Chem. Phys.*, **76**, 617 (1982).
49. R. K. Heenan and L. S. Bartell, *J. Chem. Phys.*, **78**, 1270 (1983).
50. W. F. Murphy, *J. Raman Spectrosc.*, **23**, 413–413 (1992).
51. N. L. Allinger, R. S. Grev, B. F. Yates, and H. F. Schaefer III, *J. Am. Chem. Soc.*, **112**, 114 (1990).
52. P. Jungwirth and R. Zahradnik, *Chem. Phys. Lett.*, **212**, 211–217 (1993).
53. R. A. Scott and H. A. Scheraga, *J. Chem. Phys.*, **44**, 3054 (1966).

54. P. B. Woller and E. W. Garbisch Jr., *J. Am. Chem. Soc.*, **54**, 9310 (1972).
55. W. L. Jorgensen, J. D. Madura, and C. J. Swenson, Optimized intermolecular potential functions for liquid hydrocarbons, *J. Am. Chem. Soc.*, **106**, 6638–6646 (1984).
56. A. Banon, F. Serrano Adan, and J. Santamaria, *J. Chem. Phys.*, **83**, 297 (1985).
57. W. L. Jorgensen and J. K. Buckner, *J. Phys. Chem.*, **91**, 6083 (1987).
58. E. Enciso, J. Alonso, and N. G. Almarza, *J. Chem. Phys.*, **90**, 413–421 (1989).
59. C. Chotia, *Annu. Rev. Biochem.*, **53**, 537–572 (1984).
60. C. L. Brooks III, M. Karplus, and B. M. Petitt, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Wiley, New York, 1988.
61. J. Tirado-Reves and W. L. Jorgensen, *Biochemistry*, **30**, 3864–3871 (1991).
62. I. L. Karle, J. Flippen-Andersen, K. Uma, and P. Balaram, *Curr. Sci.*, **59**, 875–885 (1990).
63. A. Bavioso, E. Benedetti, B. DiBlasio, V. Pavone, C. Pedone, C. Toniolo, and G. M. Bonora, *Proc. Natl. Acad. Sci. USA*, **83**, 1988–1992 (1986).
64. M. L. McKee and M. Page, In K. B. Lipkowitz and D. B. Boyd, Eds., *Reviews in Computational Chemistry*, Vol. IV, VCH, New York, 1993.
65. H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, MA, 1980.